UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/966,819	09/28/2001	Patrick Duncan Jenny	08204/000S097-US0/10.032	3058

38878 7590 07/20/2009

F5 Networks, Inc.
c/o DARBY & DARBY P.C.
P.O. BOX 770
Church Street Station
NEW YORK, NY 10008-0770

EXAMINER	
MEUCCI, MICHAEL D	
ART UNIT	PAPER NUMBER
2442	

MAIL DATE	DELIVERY MODE
07/20/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	09/966,819	JENNY ET AL.	
	Examiner	Art Unit	
	MICHAEL D. MEUCCI	2442	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *15 December 2008*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-25,27 and 28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-25,27 and 28* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *28 September 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. This action is in response to the request for reconsideration filed 15 December 2008.

2. Claims 1-25, 27, and 28 remain pending.

Claim Rejections - 35 USC § 103

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

4. Claims 8-10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wu et al. (U.S. 6,370,620 B1) hereinafter referred to as Wu, in view of Scharber (U.S. 6,542,964 B1) and Lamburt et al. (U.S. 6,374,241 B1) hereinafter referred to as Lamburt.

a. As per claim 8, Wu teaches: determining a frequency of requests for a plurality of different static content set, wherein the content set includes the requested content (lines 52-54 of column 6 and lines 47-58 of column 1 – which clearly show multiple objects available for requests, i.e. *different static content*); if the frequency of requests for the static content exceeds a threshold, forwarding the request for the static content to the cache (lines 43-59 of column 6); wherein the content is obtained when unavailable in the cache by generating another request for the content and forwards the request to another cache determined by

hashing an identifier associated with the static content if the frequency of requests for static content is below the threshold (line 48 of column 4 through line 8 of column 5 and lines 21-28 of column 6).

Wu does not explicitly teach: determining at least one type of the requested content based on a determination of information included within the request; and a first cache that employs a hot list for access to static content that is separately cached.

Regarding determining at least one type of the requested content based on a determination of information included within the request, Scharber discloses: "By being able to recognize the content type associated with these different requests (e.g., based on the transport protocol or otherwise), ICDS 50 is able to determine which caching protocol is appropriate," (lines 40-43 of column 7). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to determine at least one type of the requested content based on a determination of information included within the request. "That is, ICDS 50 is able to make a content deterministic evaluation of the appropriate cache protocol to be used," (lines 43-45 of column 7 in Scharber). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to determine at least one type of the requested content based on a determination of information included within the request in the system as taught by Wu.

The combination of Wu and Scharber does not explicitly teach: a first cache that employs a hot list for access to static content that is separately

cached. However, Lamburt discloses: "The Data Query Cache 850, in this embodiment, generally includes a "hot" and "cold" cache," (lines 36-37 of column 27) and the hot and cold caches inherently contain a list of objects in them. It would have been obvious to one of ordinary skill in that art at the time of the applicant's invention to have a first cache that employs a hot list for access to static content that is separately cached. "In this embodiment, the caching technique implemented is the LRU (Least Recently Used) policy by which elements of the cache are selected for replacement in accordance with time from last use. These and other policies are generally known to those skilled in the art. Generally, the "hot" cache may include the most recently used items and the cold cache the remaining items," (lines 37-43 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have a first cache that employs a hot list for access to static content that is separately cached in the system as taught by Wu.

b. As per claim 9, Wu teaches: hashing the identifier associated with the content to obtain a value and forwarding the request to a cache associated with the value when the frequency of requests for the plurality of different static content in the content set is below the threshold (line 48 of column 4 through line 8 of column 5 and line 43 of column 6 through line 3 of column 7).

c. As per claim 10, Wu teaches: another request is forwarded to the content server when the static content is unavailable from the other cache (lines 4-28 of column 6).

5. Claims 12, 14-17, 19, 21, and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Wu in view of Scharber, Lamburt, Banerjia et al. (U.S. 2001/0049818 A1) hereinafter referred to as Banerjia, and Palanca et al. (U.S. 6,216,215 B1) hereinafter referred to as Palanca.

a. As per claims 12 and 25, Wu teaches: a forwarder that receives each request for content in the system and forwards each request over the network to at least one of a content server and a caches (line 48 of column 4 through line 8 of column 5); a determination of a frequency of each requests for a static type of content (lines 52-54 of column 6 and lines 47-58 of column 1 – which clearly show multiple objects available for requests, i.e. *different static content*); the content server is coupled to the forwarder wherein the content server sends content to the client in response to each request that is forwarded to the content server and the cache is coupled to the forwarder, wherein the cache sends content to the client in response to each request that is forwarded to the cache (line 48 of column 4 through line 8 of column 5 and lines 21-28 of column 6).

Wu does not explicitly teach: wherein the forwarding of each request is based on a determination of content generation information included within each request if the requested content is dynamic or static; a plurality of caches including at least one hot cache; wherein the hot cache is based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache; and a determination if a request

for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content.

Scharber discloses: "By being able to recognize the content type associated with these different requests (e.g., based on the transport protocol or otherwise), ICDS 50 is able to determine which caching protocol is appropriate," (lines 40-43 of column 7). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have the forwarding of each request based on a determination of content generation information included within each request if the requested content is dynamic or static. "That is, ICDS 50 is able to make a content deterministic evaluation of the appropriate cache protocol to be used," (lines 43-45 of column 7 in Scharber). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have the forwarding of each request based on a determination of content generation information included within each request if the requested content is dynamic or static in the system as taught by Wu.

The combination of Wu and Scharber does not explicitly teach: a plurality of caches including at least one hot cache; wherein the hot cache is based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache; and a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content. However, Lamburt discloses: "The Data Query

Cache 850, in this embodiment, generally includes a "hot" and "cold" cache," (lines 36-37 of column 27). It would have been obvious to one of ordinary skill in that art at the time of the applicant's invention to have a plurality of caches including at least one hot cache. "In this embodiment, the caching technique implemented is the LRU (Least Recently Used) policy by which elements of the cache are selected for replacement in accordance with time from last use. These and other policies are generally known to those skilled in the art. Generally, the "hot" cache may include the most recently used items and the cold cache the remaining items," (lines 37-43 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have a plurality of caches including at least one hot cache in the system as taught by the combination of Wu and Scharber.

The combination of Wu, Scharber, and Lamburt does not explicitly teach: wherein the hot cache is based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache; and a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content. Lamburt teaches the hot cache including the LRU (Least Recently Used) policy, but does not explicitly teach higher frequency usage for the hot cache. However, Banerjia discloses: “By tracking the execution frequency of each translation, the code cache can obtain canonical information about which translations are executed the most frequently. The code cache can then user this information, along with a

“hot threshold" to classify all translations into a plurality of different sets, based on their frequency of execution,” (paragraph [0026] on page 2). It would have been obvious for one of ordinary skill in the art at the time of the applicant’s invention to have the hot cache based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache. “However, it should be clear to one skilled in the art that two or more different thresholds could be provided in order to create three or more separate partitions in the code cache, with each partition storing translations in a different non-overlapping range of execution frequencies," (paragraph [0026] on page 2 of Banerjia). It is for this reason that one of ordinary skill in the art at the time of the applicant’s invention would have been motivated to have the hot cache based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache in the system as taught by the combination of Wu, Scharber, and Lamburt.

The combination of Wu, Scharber, Lamburt, and Banerjia does not explicitly teach: a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content. However, Palanca discloses: For senior load retirement from the L1CC 250, the L1CC 250 asserts the write-back data valid signal upon L1 cache hit or upon L1 buffer allocation (if there is a L1 cache miss), and not upon the return of the requested data. On the other hand, for a senior load L1 cache miss, the L1CC masks (i.e., clears) the write-back data valid signal upon the return of the requested data,”

(lines 1-7 of column 10) wherein the write-back data valid signal denotes that the request is coming from a cache that it had previously forwarded a request (on a cache miss). It would have been obvious for one of ordinary skill in the art at the time of the applicant's invention to have a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content. "The write-back masking also avoids contention on the writeback bus with another instruction. This is implemented by masking (i.e., clearing) the write-back data valid signal to the re-order buffer and register file 220. The L1 cache controller 250 retires all non-senior loads by asserting the write-back data valid signal when the requested data is available," (lines 7-13 of column 10 in Palanca). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content in the system as taught by the combination of Wu, Scharber, Lamburt, and Banerjia.

b. As per claims 14-15, Wu teaches: the forwarder is coupled to the content server over a wide area network/local area network; and the forwarder is coupled to the content server over a communications medium (line 47 of column 2 through line 14 of column 3 and Fig. 1).

c. As per claim 16, Wu teaches: the information includes at least one of where the request is generated, the frequency of requests for the content, and the nature of the content requested (lines 16-28 of column 7).

d. As per claim 17, Wu teaches: the forwarder is structured to forward requests to the content server when the information indicated that the request is generated by the regular cache (lines 20-28 of column 6).

e. As per claim 19, Wu teaches: forwarding requests when not found in primary cache (lines 20-28 of column 6).

The above-described combination of Wu, Scharber, Lamburt, Banerjia, and Palanca does not explicitly teach: the forwarder is further structured to forward requests to the regular cache when the information indicates that the request is generated by the hot cache. However, Lamburt discloses: "It should generally be noted that in this particular embodiment, the "hot" cache is implemented as storing the data in random access memory," (lines 48-50 of column 27). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to utilize the "hot" cache as the primary cache in the system of Wu. "This may be distinguished from the storage medium associated with the "cold" cache representing those items which are determined, in accordance with caching policies such as the LRU, to be least likely to be accessed when compared with the items in the hot cache which are determined to be more likely to be accessed," (lines 50-56 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to utilize the "hot" cache of Lamburt as the

primary cache in Wu, and to forward the request to the regular cache when the information indicates that the request is generated by the hot cache.

f. As per claim 21, Wu teaches: the server uses a hash table to calculate the number of requests for the content (lines 44-61 of column 1 and line 48 of column 4 through line 8 of column 5).

6. Claims 1, 24, and 27-28 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout (U.S. 5,566,349) in view of Lamburt, Scharber, Banerjia, and Jordan et al. (U.S. 2002/0026560 A1) hereinafter referred to as Jordan.

As per claims 1, 24, and 27-28, Trout teaches: receiving a request for content from a client (lines 35-37 of column 4 and lines 19-28 of column 42) and determining at least one type of the requested content based on information included within the request (lines 39-42 of column 11, lines 61-64 of column 27, and lines 31-34 of column 28); when the type of the requested content is dynamic, forwarding the request to a content server that enables access to the dynamic content (lines 10-11 of column 12); and when the type of the requested content is static, forwarding the request to a cache that enables access to the static content (lines 11-12 of column 12).

Trout does not explicitly teach: a plurality of caches including at least one hot cache that is based at least in part on a higher frequency of request over a period of time; determining at least one type of the requested content based on a determination of the request; and wherein the plurality of caches is organized in a hierarchy and wherein a higher level cache in the hierarchy is associated with a

higher frequency of requests for static content than a lower frequency of requests for static content associated with a lower level cache, and wherein forwarding the request over the network to the plurality of caches that enable access to the static content further comprises recursively forwarding requests, generated from different caches in the hierarch based on the received request and receipt of one of the recursively forwarded requests, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with the hot cache.

Regarding having a plurality of caches including at least one hot cache, Lamburt discloses: "The Data Query Cache 850, in this embodiment, generally includes a "hot" and "cold" cache," (lines 36-37 of column 27). It would have been obvious to one of ordinary skill in that art at the time of the applicant's invention to have a plurality of caches including at least one hot cache. "In this embodiment, the caching technique implemented is the LRU (Least Recently Used) policy by which elements of the cache are selected for replacement in accordance with time from last use. These and other policies are generally known to those skilled in the art. Generally, the "hot" cache may include the most recently used items and the cold cache the remaining items," (lines 37-43 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have a plurality of caches including at least one hot cache in the system as taught by Trout.

The combination of Trout and Lamburt does not explicitly teach: cache that is based at least in part on a higher frequency of request over a period of

time; determining at least one type of the requested content based on a determination of the request; and wherein the plurality of caches is organized in a hierarchy and wherein a higher level cache in the hierarchy is associated with a higher frequency of requests for static content than a lower frequency of requests for static content associated with a lower level cache, and wherein forwarding the request over the network to the plurality of caches that enable access to the static content further comprises recursively forwarding requests, generated from different caches in the hierarch based on the received request and receipt of one of the recursively forwarded requests, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with the hot cache. However, regarding determining at least one type of the requested content based on a determination of the request: Scharber discloses: "By being able to recognize the content type associated with these different requests (e.g., based on the transport protocol or otherwise), ICDS 50 is able to determine which caching protocol is appropriate," (lines 40-43 of column 7). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to determine at least one type of the requested content based on a determination of the request. "That is, ICDS 50 is able to make a content deterministic evaluation of the appropriate cache protocol to be used," (lines 43-45 of column 7 in Scharber). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to determine at least one type of the requested content based on a determination of the request in the system as taught by the combination of Trout and Lamburt.

The combination of Trout, Lamburt, and Scharber does not explicitly teach: higher frequency usage for the hot cache, wherein the plurality of caches is organized in a hierarchy and wherein a higher level cache in the hierarchy is associated with a higher frequency of requests for static content than a lower frequency of requests for static content associated with a lower level cache, and wherein forwarding the request over the network to the plurality of caches that enable access to the static content further comprises recursively forwarding requests, generated from different caches in the hierarch based on the received request and receipt of one of the recursively forwarded requests, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with the hot cache. Lamburt teaches the hot cache including the LRU (Least Recently Used) policy, but does not explicitly teach higher frequency usage for the hot cache. However, Banerjia discloses: “By tracking the execution frequency of each translation, the code cache can obtain canonical information about which translations are executed the most frequently. The code cache can then user this information, along with a “hot threshold" to classify all translations into a plurality of different sets, based on their frequency of execution,” (paragraph [0026] on page 2). It would have been obvious for one of ordinary skill in the art at the time of the applicant’s invention to have the hot cache based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache. “However, it should be clear to one skilled in the art that two or more different thresholds could be provided in order to create three or more separate partitions in the code cache,

with each partition storing translations in a different non-overlapping range of execution frequencies," (paragraph [0026] on page 2 of Banerjia). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have the hot cache based at least in part on the request for static content with a higher frequency greater than a lower frequency associated with a lower level cache in the system as taught by the combination of Trout, Lamburt, and Scharber.

The combination of Trout, Lamburt, Scharber, and Banerjia does not explicitly teach: wherein the plurality of caches is organized in a hierarchy and wherein a higher level cache in the hierarchy is associated with a higher frequency of requests for static content than a lower frequency of requests for static content associated with a lower level cache, and wherein forwarding the request over the network to the plurality of caches that enable access to the static content further comprises recursively forwarding requests, generated from different caches in the hierarch based on the received request and receipt of one of the recursively forwarded requests, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with the hot cache. However, Jordan discloses: "The present invention also includes features for periodically monitoring the load condition on and the forwarding frequency to the owning cache server; and proactively shifting one or more subsequent forwarded requests for the cached object from the owning cache server to one or more of the cooperating cache servers, in response to the monitoring. Alternatively, the shifting step further includes the step of checking the load condition and

forwarding frequency, in response to the receipt of a forwarded request," (paragraph [0013] on page 2) wherein the proactive shifting of one or more subsequent forwarded requests teaches the recursive nature of the system. It would have been obvious for one of ordinary skill in the art at the time of the applicant's invention to have caches organized in a hierarchy and forward the request over the network to the plurality of caches that enable access to the static content further comprise recursively forwarding requests, generated from different caches in the hierarch based on the received request, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with the hot cache. "Depending on the load condition 10212 and forwarding frequency 1011 of requests for p 10101 on server B, the load monitor may forward the request to server B, asking it to send a copy of object p to server C. Or, if server B is currently overloaded or is trending as such, the load monitor might shift the forwarded request by finding an underloaded (or less loaded) server to serve as a new (or shared as in B, A 10121) owner of object p. The request is then forwarded to the new (or shared e.g., A) owning server for the object," (paragraph [0032] on page 4 of Jordan). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have caches organized in a hierarchy and forward the request over the network to the plurality of caches that enable access to the static content further comprise recursively forwarding requests, generated from different caches in the hierarch based on the received request, through the hierarchy until a frequency of the request for static content exceeds a threshold associated with

the hot cache in the system as taught by the combination Trout, Lamburt, Scharber, and Banerjia.

7. Claim 2 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout, Lamburt, Scharber, Banerjia, and Jordan further in view of Factor et al. (U.S. 6,094,706 B1) hereinafter referred to as Factor.

The combination of Trout, Lamburt, Scharber, Banerjia, and Jordan described above does not explicitly teach: the hot cache caches static content when a frequency of requests for the static context exceeds a threshold. However, Factor discloses: "Once a particular component has been accessed more than a threshold number of times, new pathnames that contain this component may be added to the °cache," (lines 52-54 of column 11). It would have been obvious to one of ordinary skill in that art to have the hot cache cache static content when a frequency of requests for the static context exceeds a threshold. "This component may be added to the cache under the assumption that the new pathnames will also be accessed frequently," (lines 54-56 of column 11 in Factor). It is for this reason that one of ordinary skill in that art at the time of the applicant's invention would have been motivated to have the hot cache cache static content when a frequency of requests for the static context exceeds a threshold in the system as taught by the combination of Trout, Lamburt, Scharber, Banerjia, and Jordan.

8. Claim 3 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout, Lamburt, Scharber, Banerjia, and Jordan, further in view of Guenthner et al. (U.S. 5,590,301) hereinafter referred to as Guenthner.

The combination of Trout, Lamburt, Scharber, Banerjia, and Jordan does not explicitly teach: when the static content is unavailable in the hot cache, forwarding the request to another cache in the plurality of caches. However, Guenthner discloses: "an internal address, including a cluster number, is sent to the address translator 18 as a request from the primary cache directed to the secondary cache 7 (which, of course, will forward the request to main memory if the requested information is not resident in the secondary cache at the time of the request)," (lines 21- 26 of column 7). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to forward the request to another cache when the static content is unavailable in the hot cache. "Register 15 is merely a convenient representation of address interface circuitry in the primary cache of the CPU 11 by which an address generated by the CPU 11 may be transmitted, transformed in the address translator 18, as a request to the secondary cache 7. This condition occurs when information required by the CPU 11 is not resident in at least one of the primary caches of the CPUs 11, 12, 13, 14 on the multiprocessor board 1. (Those skilled in the art will understand that, in many such multiprocessor configurations, it is possible for one CPU to "siphon" information from another CPU's primary cache)," (lines 18-28 of column 4 in Guenthner). It is for this reason that one of ordinary skill in that art at the time of the applicant's invention would have been motivated to forward the request to

another cache when the static content is unavailable in the hot cache in the system as taught by the combination of Trout, Lamburt, Scharber, Banerjia, and Jordan.

9. Claim 4 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout, Lamburt, Scharber, Banerjia, and Jordan, further in view of McCanne (U.S. 6,785,704 B1).

The combination of Trout, Lamburt, Scharber, Banerjia, and Jordan does not explicitly teach: when the static content is unavailable from any one of the plurality of caches, forwarding the request to the content server that enables access to the static content. However, McCanne discloses: "the cache serves the request, if it can, or forwards the request to the content server and then serves the client the content returned from the content server." (lines 63-65 of column 3). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to forward the request to the content server that enables access to the static content when the static content is unavailable from any one of the plurality of caches. "Caching can be either transparent or nontransparent. With transparent caching, the client makes a request of the content server and the network infrastructure intercepts the request if the cache can serve the request. With nontransparent caching, the client makes the request of the cache (or more precisely, of a network node to which the cache is attached) and the cache serves the request, if it can, or forwards the request to the content server and then serves the client the content returned from the content server," (lines

57-65 of column 3 in McCanne). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to forward the request to the content server that enables access to the static content when the static content is unavailable from any one of the plurality of caches in the system as taught by the combination of Trout, Lamburt, Scharber, Banerjia, and Jordan.

10. Claim 5 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout, Lamburt, Scharber, Banerjia, and Jordan, further in view of Kimura et al. (U.S. 6,415,359 B1) hereinafter referred to as Kimura.

The combination of Trout, Lamburt, Scharber, Banerjia, and Jordan does not explicitly teach: examining the request for an extension indicating that a process is performed in response to the request, wherein the process includes at least one of an application program and a script. However, Kimura discloses: "in the case of creating a new file in the portable information processing terminal device 10 in response to a request from another information processing device, the file management unit 102 first checks the attribute information (an extension and a file name, or other ID information indicating a file type, etc.) of that file which is attached to the creation request (step $71), and judges whether it is a file that should be stored into the cache 17 or not (step $72)o An application program file that is executable on the portable information processing terminal device 10 or a file that can be processed by that application has a high probability of being accessed in the disk access prohibited state during the

battery driven mode so that such a file will be judged as a file that should be stored into the cache 17," (lines 49-62 of column 13). It would have been obvious to one of ordinary skill in that art at the time of the applicant's invention to examine the request for an extension indicating that a process is performed in response to the request, wherein the process includes at least one of an application program and a script. "An application program file that is executable on the portable information processing terminal device 10 or a file that can be processed by that application has a high probability of being accessed in the disk access prohibited state during the battery driven mode so that such a file will be judged as a file that should be stored into the cache 17. In the case where the judgment cannot be made, it is also possible to inquire the user as to whether it is a file that should be stored into the cache 17 or not," (lines 57-65 of column 13 in Kimura). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to examine the request for an extension indicating that a process is performed in response to the request, wherein the process includes at least one of an application program and a script in the system as taught by the combination of Trout, Lamburt, Scharber, Banerjia, and Jordan.

11. Claims 6-7 rejected under 35 U.S.C. 103(a) as being unpatentable over Trout, Lamburt, Scharber, Banerjia, and Jordan, further in view of Dujari (U.S. 6,233,606 B1).

The combination of Trout, Lamburt, Scharber, Banerjia, and Jordan does not explicitly teach: the content includes information associated with a plurality of resource identifiers; and the resource identifiers are uniform resource locators (URLs). However, Dujari discloses: "the content can be indexed by a unique lookup key, such as a Uniform Resource Identifier (URI), a compact string of characters for identifying an abstract or physical resource. Examples of URIs include URLs (Uniform Resource Locators), URNs (Uniform Resource Names), and other standard namespaces," (lines 28-33 of column 1). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have the content include information associated with a plurality of resource identifiers; and have the resource identifiers as uniform resource locators (URLs). "A URI may be used as the lookup key to a cache, as can other names, such as a globally unique identifier (GUID)," (lines 33-35 of column 1 in Dujari)o It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have the content include information associated with a plurality of resource identifiers; and have the resource identifiers as uniform resource locators (URLs) in the system as taught by the combination of Trout, Lamburt, Scharber, Banerjia, and Jordan.

12. Claim 11 rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Scharber and Lamburt as applied to claim 10 above, in view of Cohen et al. (U.S. 6,330,561 B1) hereinafter referred to as Cohen.

Wu teaches: another request is forwarded to the content server when the content is unavailable from the other cache (lines 4-28 of column 6).

The combination of Wu, Scharber, and Lamburt does not explicitly teach: the content server forwards the other request for content to an additional cache. However, Cohen discloses: "Then the proxy server would forward a request for validation with respect to the client requested resource and a request for validation with regard to one or more additional resources in the proxy cache that were from the same resource server," (lines 30-34 of column 2). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have the content server forward the other request for content to an additional cache. "This approach is a benefit to the proxy cache in the sense that it helps the proxy cache determine the validity of certain of its contents at an earlier time," (lines 38-40 of column 2 in Cohen). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have the content server forward the other request for content to an additional cache in the system as taught by the combination of Wu, Scharber, and Lamburt.

13. Claim 13 rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Scharber, Lamburt, Banerjia, and Palanca as applied to claim 12 above, in view of Cohen and Sharma (U.S. 6,591,341 B1).

Wu teaches: a regular cache and forwarding requests if content is not found in cache.

The combination of Wu, Scharber, and Lamburt, Banerjia, and Palanca does not explicitly teach: a hot cache and an additional cache, wherein the hot cache, the regular cache, and the additional cache are arranged in a hierarchical order for receiving each forwarded request for content from the forwarder.

However, Lamburt discloses: "It should generally be noted that in this particular embodiment, the "hot" cache is implemented as storing the data in random access memory," (lines 48-50 of column 27). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have a hot cache. "This may be distinguished from the storage medium associated with the "cold" cache representing those items which are determined, in accordance with caching policies such as the LRU, to be least likely to be accessed when compared with the items in the hot cache which are determined to be more likely to be accessed," (lines 50-56 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have a hot cache in the system as taught by Wu and Scharber.

Regarding having an additional cache system, Cohen discloses: "Then the proxy server would forward a request for validation with respect to the client requested resource and a request for validation with regard to one or more additional resources in the proxy cache that were from the same resource server," (lines 30-34 of column 2). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have an additional cache. "This approach is a benefit to the proxy cache in the sense that it helps the proxy

cache determine the validity of certain of its contents at an earlier time," (lines 38-40 of column 2 in Cohen). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have an additional cache in the system as taught by Wu, Scharber, and Lamburt.

The combination of Wu, Scharber, Lamburt, Banerjia, Palanca, and Cohen does not teach: to arrange the cache in a hierarchical order for receiving each forwarded request for content from the forwarder. However, Sharma discloses: "If the request was a cache miss in the second data array, the request may be forwarded to another level of memory hierarchy, such as another cache or a system memory (lines 32-35 of column 5 and Fig. 5). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to arrange the cache in a hierarchical order for receiving each forwarded request for content from the forwarder. "In either case, when it was determined that there was a cache miss in the first data array, the one or more instructions that were tentatively processed may be replayed," (lines 35-37 of column 5 in Sharma). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to arrange the cache in a hierarchical order for receiving each forwarded request for content from the forwarder in the system as taught by the combination of Wu, Scharber, Lamburt, Banerjia, Palanca, and Cohen.

14. Claim 18 rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Scharber, Lamburt, Banerjia, and Palanca as applied to claim 16 above, in view of Factor.

The combination of Wu, Scharber, Lamburt, Banerjia, and Palanca does not explicitly teach: the forwarder is further structured to forward requests to the hot cache when the information indicates that the rate of requests exceeds a threshold. However, Lamburt discloses: "It should generally be noted that in this particular embodiment, the "hot" cache is implemented as storing the data in random access memory," (lines 48-50 of column 27). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to forward requests to a hot cache. "This may be distinguished from the storage medium associated with the "cold" cache representing those items which are determined, in accordance with caching policies such as the LRU, to be least likely to be accessed when compared with the items in the hot cache which are determined to be more likely to be accessed," (lines 50-56 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to forward requests to a hot cache in the system as taught by Wu and Scharber. Factor discloses: "Once a particular component has been accessed more than a threshold number of times, new pathnames that contain this component may be added to the cache," (lines 52-54 of column 11). It would have been obvious to one of ordinary skill in that art to forward requests to a hot cache when the rate of requests exceeds a threshold. "This component may be added to the cache under the assumption that the new

pathnames will also be accessed frequently," (lines 54-56 of column 11 in Factor). It is for this reason that one of ordinary skill in that art at the time of the applicant's invention would have been motivated to forward requests to a hot cache when the rate of requests exceeds a threshold in the system as taught by the combination of Wu, Scharber, Lamburt, Banerjia, and Palanca.

15. Claim 20 rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Scharber, Lamburt, Banerjia, and Palanca as applied to claim 12 above, in view of Sharma.

The combination of Wu, Scharber, Lamburt, Banerjia, and Palanca does not explicitly teach: the hot cache and the regular cache are located on the same device. However, Lamburt discloses: "The Data Query Cache 850, in this embodiment, generally includes a "hot" and "cold" cache," (lines 36-37 of column 27). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to utilize the "hot" cache as the primary cache in the system of Wu. "This may be distinguished from the storage medium associated with the "cold" cache representing those items which are determined, in accordance with caching policies such as the LRU, to be least likely to be accessed when compared with the items in the hot cache which are determined to be more likely to be accessed," (lines 50-56 of column 27 in Lamburt). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to utilize the "hot" cache of Lamburt as the primary cache in Wu and Scharber.

With the addition of the section of this section of Lamburt, the combination of Wu, Scharber, Lamburt, Banerjia, and Palanca still does not teach: having multiple caches on the same device. However, Sharma discloses: "Many computer, systems use multiple levels of caches to cache data from a memory device. For example, a computer system may have a level one cache (L1) and a larger level two cache (L2), in addition to an even larger RAM memory," (lines 14-17 of column 1). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have multiple caches on the same device. "The L1 cache typically contains a copy of information that was previously loaded from RAM by the processor, and the L2 cache typically contains both a copy of information in the L1 cache and other information that had been loaded from RAM by the processor less recently than the information in the L1 cache," (lines 18-24 of column 1 in Sharma). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have multiple caches on the same device in the system as taught by the combination of Wu, Scharber, Lamburt, Banerjia, and Palanca.

16. Claims 22-23 rejected under 35 U.S.C. 103(a) as being unpatentable over Wu, Scharber, Lamburt, Banerjia, and Palanca as applied to claim 12 above, in view of Dujari.

The combination of Wu, Scharber, Lamburt, Banerjia, and Palanca does not explicitly teach: the content includes information associated with a plurality of resource identifiers; and the resource identifiers are uniform resource locators

(URLs). However, Dujari discloses: "the content can be indexed by a unique lookup key, such as a Uniform Resource Identifier (URI), a compact string of characters for identifying an abstract or physical resource. Examples of URIs include URLs (Uniform Resource Locators), URNs (Uniform Resource Names), and other standard namespaces," (lines 28-33 of column 1). It would have been obvious to one of ordinary skill in the art at the time of the applicant's invention to have the content include information associated with a plurality of resource identifiers; and have the resource identifiers as uniform resource locators (URLs). "A URI may be used as the lookup key to a cache, as can other names, such as a globally unique identifier (GUID)," (lines 33-35 of column 1 in Dujari). It is for this reason that one of ordinary skill in the art at the time of the applicant's invention would have been motivated to have the content include information associated with a plurality of resource identifiers; and have the resource identifiers as uniform resource locators (URLs) in the system as taught by the combination of Wu, Scharber, Lamburt, Banerjia, and Palanca.

Response to Arguments

17. Applicant's arguments filed 15 December 2008 have been fully considered but they are not persuasive.

18. (A) Regarding claim 8, the applicant contends that Wu does not teach determining a frequency of requests summed from request for all content of a

plurality of different static content in a content set. The examiner respectfully disagrees.

As to point (A), the applicant argues that Wu does not teach one "reference count" referring to a plurality of objects. The examiner points out that the applicant's amendment which includes: "determining a frequency of requests summed from requests for all content of a plurality of different static content in a content set" does not distinguish over the cited art because the frequency is attained by merely counting the number of requests for a particular group of data. While Wu teaches counting the number of requests for a particular object (lines 52-54 of column 6), is it understood that any number of different groupings may be used when determining how many requests are made. This is a design choice that does not affect the way the system works, but merely changes the results, which are inconsequential to the system by nature. As such, the rejection remains proper and is maintained by the examiner.

19. (B) Regarding claim 10, the applicant contends that Wu does not teach: "a third request" in the limitation "when the content is unavailable from the second cache, a third request for the content is forwarded over the network to a content server." The examiner respectfully disagrees.

As to point (B), the applicant argues that Wu teaches a single request that is initially received and then re-directed or returned. The examiner points out that Wu first teaches a second request: (lines 14-17 of column 6) "Once it decides to redirect the request, block 703, the web cache server 4 sends the request back

to the browser 9 along with an IP address corresponding to the suggested sibling web cache server 4." It is clear from this recitation that the browser will send a second request for the contents, this time to the sibling web cache server. It is well understood in the art that this type of forwarding may be done across multiple servers with multiple re-directions. Wu discloses: (lines 43-46 of column 6), "Upon receiving a request for a non-assigned-partition object 103, the front-end router 803 decides whether to forward it to the partition owner cluster 801 or to service it within its own cluster 801." It is clear that the partition owner cluster, which receives the second request, may at this point undergo the process mentioned above, with reference to lines 14-17 of column 6, and redirect the request to a sibling web cache server. One of ordinary skill in that art would also recognize that these steps may be repeated for any number of redirections and may be recursive in nature. As such, the rejection remains proper and is maintained by the examiner.

20. (C) Regarding claim 12, the applicant contends that Palanca does not teach wherein the forwarding of each request is based ...on a determination if a request for the content received at the forwarder is coming from a cache in the plurality of caches to which the forwarder previously forwarded a prior request over the network for the content. The examiner respectfully disagrees.

As to point (C), the applicant argues that Palanca pertains to an art that is unrelated to that of Wu, Scharber, and Banerjia. The examiner points out that the write-back data valid signal described in Palanca can be utilized by any type

of system and not just that of a microprocessor. The methodology can easily be transfigured to be functional in the web cache systems as is described throughout Palanca exemplified by external caches and cache subsystems. Accordingly, the rejection remains proper and is maintained by the examiner.

21. (D) Regarding claim 1, the applicant contends that Trout does not teach forwarding a request. The examiner points out that this is a conditional limitation and its inclusion is not required by the claimed invention. In any case, the Jordan reference teaches this limitation by checking the load condition and forwarding frequency, in response to the receipt of a forwarded request (paragraph [0013] on page 2). It is clear from this recitation that the provided art describes the argued limitation. As such, the rejection remains proper and is maintained by the examiner.

22. (E) The applicant's remaining arguments are directed towards subject matter described above in points (A) through (D).

Conclusion

23. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory

action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

24. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael Meucci at (571) 272-3892. The examiner can normally be reached on Monday-Friday from 9:00 AM to 6:00 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Andrew Caldwell, can be reached at (571) 272-3868. The fax phone number for this Group is 571-273-8300.

Communications via Internet e-mail regarding this application, other than those under 35 U.S.C. 132 or which otherwise require a signature, may be used by the applicant and should be addressed to [michael.meucci@uspto.gov].

All Internet e-mail communications will be made of record in the application file. PTO employees do not engage in Internet communications where there exists a possibility that sensitive information could be identified or exchanged unless the record includes a properly signed express waiver of the confidentiality requirements of 35 U.S.C. 122. This is more clearly set forth in the Interim Internet Usage Policy published in the Official Gazette of the Patent and Trademark on February 25, 1997 at 1195 OG 89.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/Andrew Caldwell/
Supervisory Patent Examiner, Art Unit 2442